

LENDING CLUB CASE STUDY

Contributors:

Sowmily Dutta and Shounak Dutta

Objective

The objective of this case study is to apply Exploratory Data Analysis (EDA) techniques to a real-world problem, gain valuable insights, and present the findings in a business-focused manner through a formal presentation.

Benefits of the case study:

- Provides insight into the practical application of EDA in addressing real-world business challenges.
- Develops a foundational understanding of risk analytics within the banking and financial services sectors.
- Demonstrates how data can be leveraged to mitigate financial losses in lending practices.
- Enhances understanding of data visualization techniques and the appropriate use of charts for real-world data analysis.

Business Understanding

The primary business objective is to minimize financial losses by accurately assessing loan applicants' risk profiles. Identifying default patterns allows the company to make informed decisions, such as approving loans for reliable applicants, adjusting terms for higher-risk individuals, or rejecting risky applications. This approach balances growth with risk management to ensure profitability while minimizing defaults.

After cleaning, the dataset contains information on 38,577 approved loan applicants with 37 columns. Key columns include `loan_amnt`, `int_rate`, `installment`, `emp_length`, `annual_inc`, `purpose`, `home_ownership`, and `loan_status`. Important financial metrics such as `revol_util`, `public_record`, `bankruptcies`, `verification_status`, `dti`, `delinq_2yrs` are also included. Some columns have missing values.

Segmented Univariate and Bivariate Analysis

The Univariate Analysis provides a detailed view of individual loan attributes, including distributions of **loan amounts**, **interest rates**, **monthly installments**, and **employment lengths**. It also examines **home ownership**, **loan purpose**, **loan terms**, and the **debt-to-income ratio**, along with **loan status**, **verification status**, and **loan grades**.

The Segmented Univariate and Bivariate analysis further explores relationships between these attributes and loan status. It investigates how **credit utilization**, **interest rates**, and **loan amounts** **correlate** with **loan outcomes**, including outlier detection for interest rates. The analysis also segments loan status by **loan grade**, **annual income**, **loan purpose**, **loan terms**, **verification status**, **debt-to-income ratios**, **home ownership**, and **bankruptcy records**. This comprehensive approach uncovers patterns and correlations that influence loan outcomes, enhancing risk assessment and decision-making.

Interest Rate Analysis and Segmentation

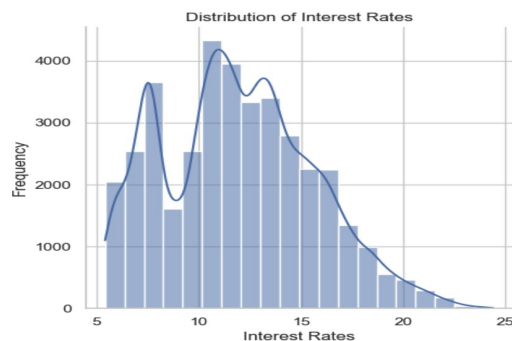


Fig 1.

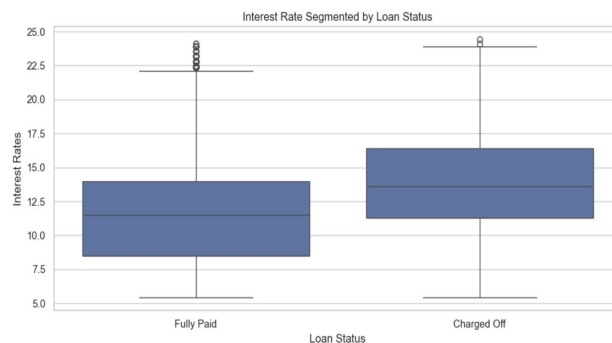


Fig 2.

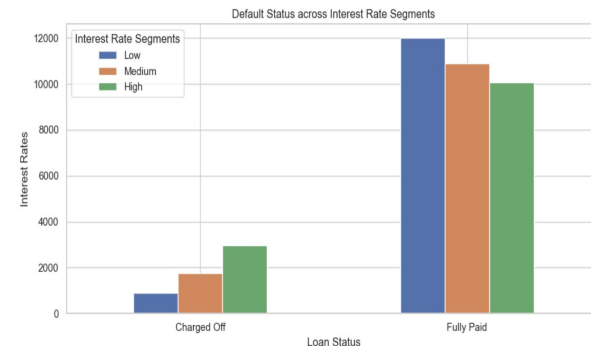


Fig 3.

Fig 1. The Interest Rate Distribution Graph shows a rightward skew, with most loans having lower interest rates. There's a peak around 10-12%, indicating that this is a common interest rate range.

Fig 2. Interest rates for fully paid loans are generally lower than for charged-off loans, indicating higher rates correlate with default risk. Charged-off loans also show greater variability in interest rates.

Fig 3. The Default Status across Interest Rate Segments Graph shows a significant increase in charged-off loans as interest rates rise, indicating a strong correlation between higher rates and default risk. While most loans with low interest rates are fully paid, a notable portion still default, emphasizing the role of other factors in loan outcomes.

Loan Amount Analysis and Segmentation

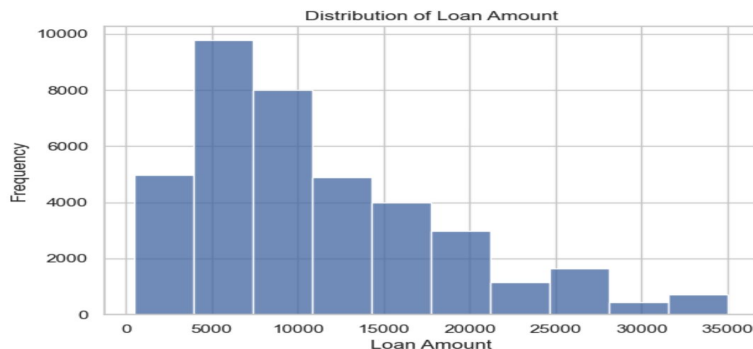


Fig 4.

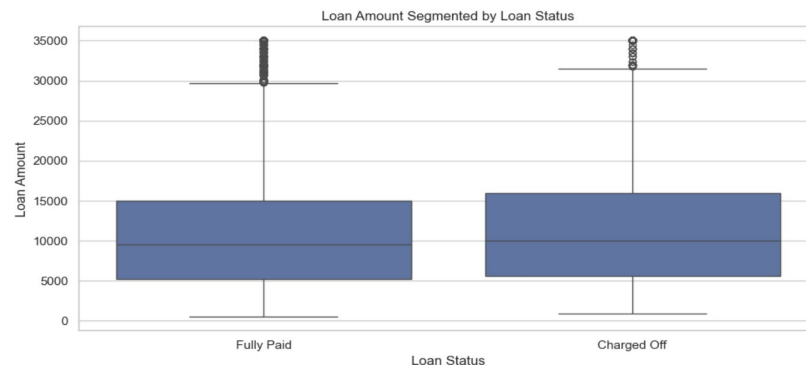


Fig 5.

Fig 4. The Loan Amount Distribution Graph shows that most loans are clustered between \$5,000 and \$15,000. There's a noticeable rightward skew, suggesting fewer loans are for much higher amounts.

Fig 5. The Loan Amount Segmented by Loan Status Graph shows that the median loan amount for fully paid loans is slightly lower than for charged-off loans, suggesting a link between larger loan amounts and higher default risk. Charged-off loans also exhibit greater variability in loan amounts.

Loan Term Analysis and Segmentation

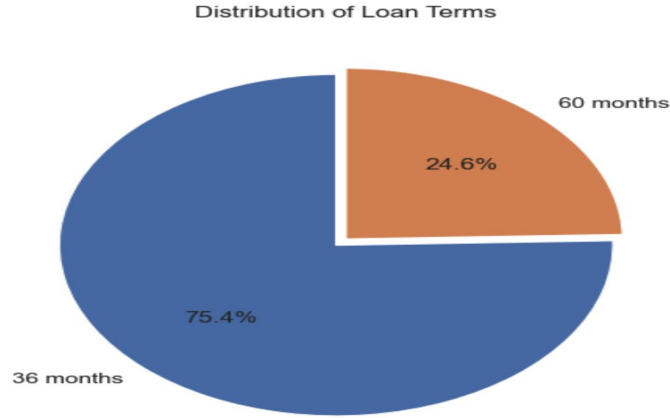


Fig 6.

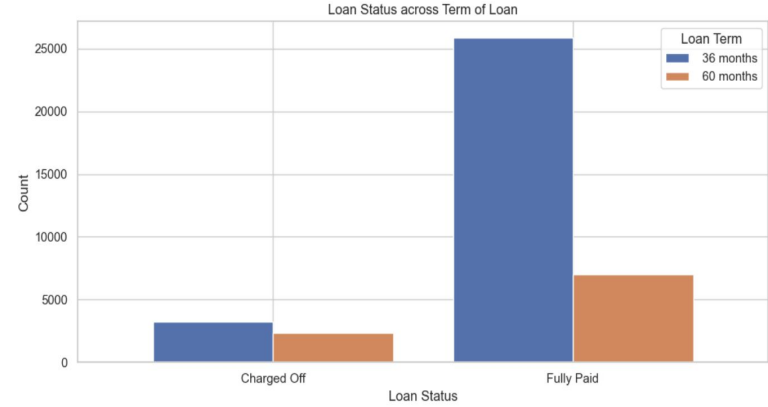


Fig 7.

Fig 6. The Loan Term Distribution Graph shows that 75.4% of loans have a 36-month term, while 24.6% have a 60-month term.

Fig 7. The Loan Status across Term of Loan Graph shows that loan term impacts default risk. Loans with a 60-month term have a significantly lower rate of charge-offs compared to those with a 36-month term.

Loan Purpose Analysis and Segmentation

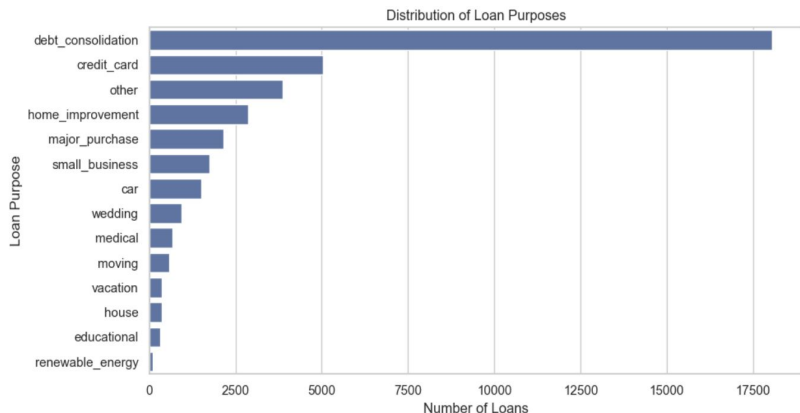


Fig 8.

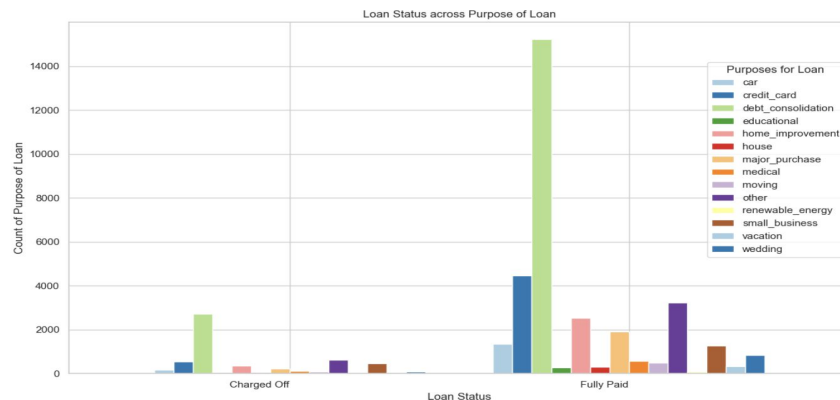


Fig 9.

Fig 8. The Loan Purpose Distribution Graph shows that debt consolidation is the most common reason for loans, followed by credit card and other purposes. Home improvement, major purchases, and small businesses are also notable categories. Less frequent but still represented purposes include weddings, medical expenses, moving, vacation, house, education, and renewable energy.

Fig 9. The Loan Status across Purpose of Loan Graph shows that the purpose of the loan affects default risk. Debt consolidation, Credit card, home improvement, and major purchases have the highest charge-off rates, while medical and education loans have the lowest. Although most loans are fully paid, categories like debt consolidation consistently show a higher proportion of charge-offs, highlighting the risk tied to these purposes.

Income Analysis and Segmentation

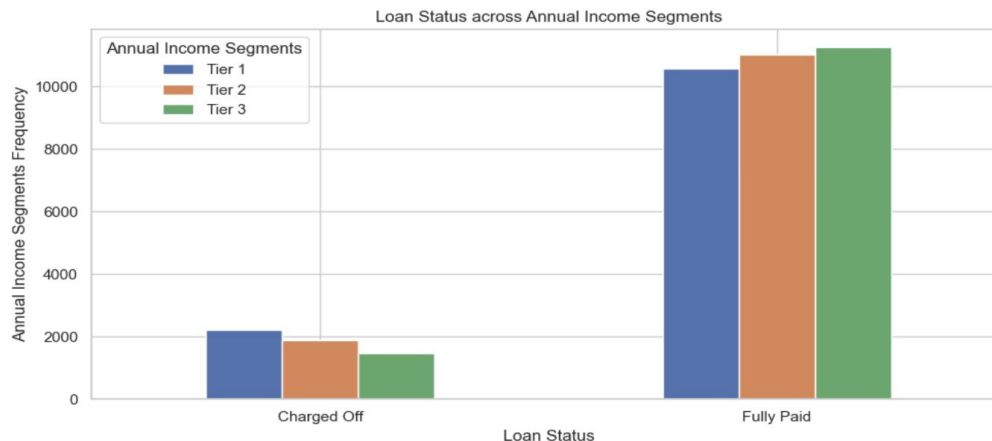


Fig 10.

Fig 10. The Loan Status across Annual Income Segments Graph shows that charged-off loans decrease as annual income increases, suggesting that borrowers in Income Tier 3 are less likely to default than those in Tier 1 and Tier 2. Here Tier 1 indicates lower income segments and Tier 3 as higher income segments.

Grade Analysis and Segmentation

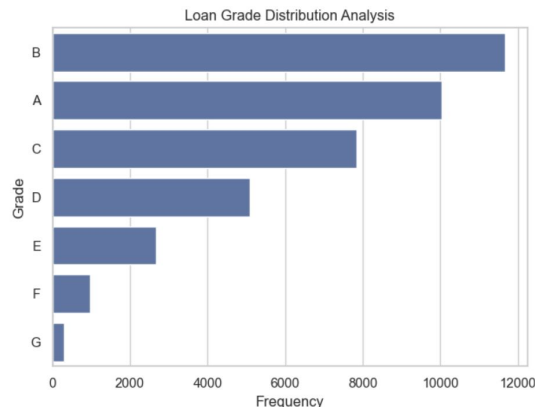


Fig 11.

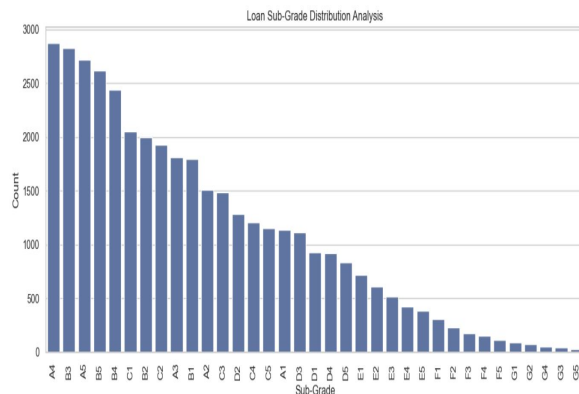


Fig 12.

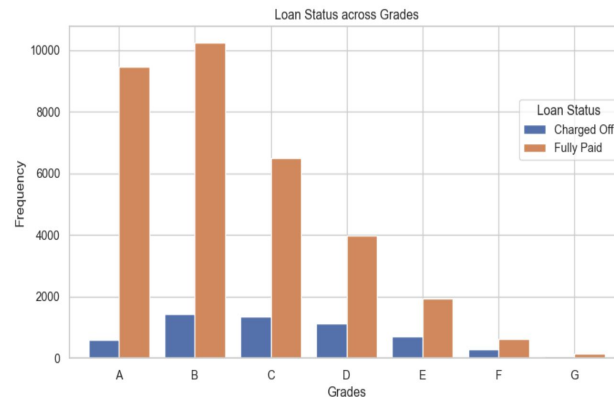


Fig 13.

Fig 11. The Loan Grade Distribution Graph shows that Grade B loans are the most common, followed by Grade A loans. The frequency of loans decreases from Grade D to G, likely due to the higher risk associated with lower grades.

Fig 12. The Loan Sub-Grade Distribution Graph is right-skewed, with more loans in the higher sub-grades (A1-A5). The most frequent sub-grade is A1, followed by B3 and A5, emphasizing the concentration of loans in the higher sub-grades. There is a noticeable decrease in the number of loans as the sub-grade moves from A1 to G5.

Fig 13. The Loan Status across Grades Graph shows that while most loans in higher grades (A-B) are fully paid, a noticeable portion still default, suggesting that factors beyond loan grade affect loan outcomes.

Verification Status Analysis and Segmentation

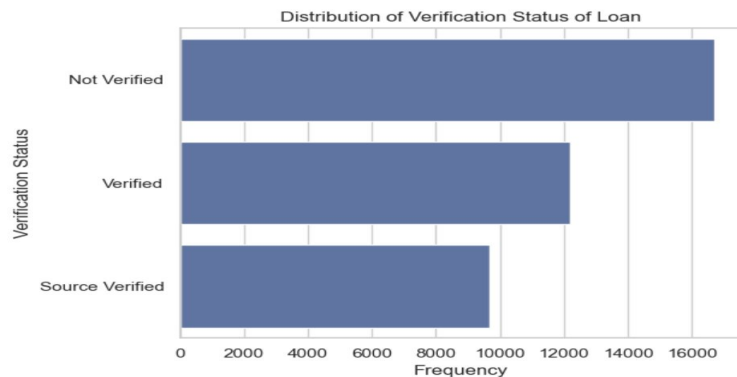


Fig 14.

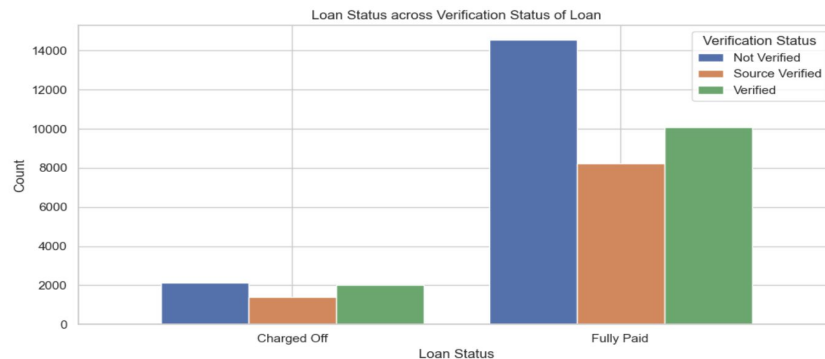


Fig 15.

Fig 14. The Verification Status Distribution Graph shows that most loans were not verified, followed by verified loans. Source verified loans represent a much smaller portion.

Fig 15. The Loan Status across Verification Status of Loan Graph shows that verification status significantly impacts default rates. Loans with no verification have the highest default rate, followed by verified loans, with source verified loans having the lowest. While most loans in all categories are fully paid, the difference in default rates underscores the importance of verification in assessing loan risk.

Debt-to-Income (DTI) Analysis and Segmentation

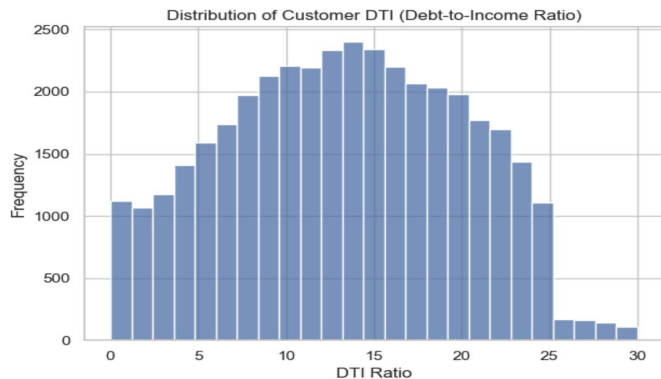


Fig 16.

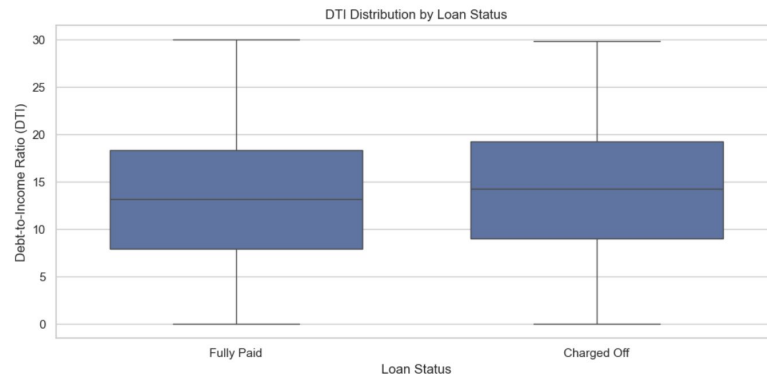


Fig 17.

Fig 16. The Customer DTI Ratio Distribution Graph is right-skewed, indicating that most borrowers have lower DTI ratios. There is a peak around the 10-15 range, suggesting this is the most common DTI ratio among borrowers.

Fig 17. The DTI Distribution by Loan Status Graph shows that the median DTI ratios for fully paid loans is slightly lower than for charged-off loans, suggesting a link between higher DTI ratios and increased default risk. Charged-off loans also exhibit greater variability in DTI ratios compared to fully paid loans.

Public Record Bankruptcies Analysis and Segmentation

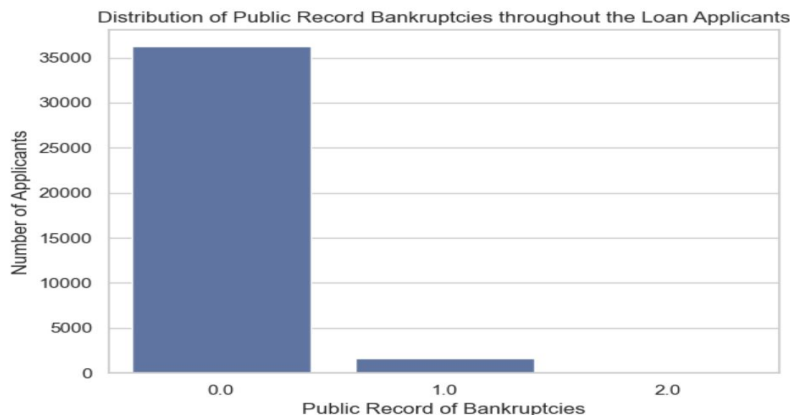


Fig 18.

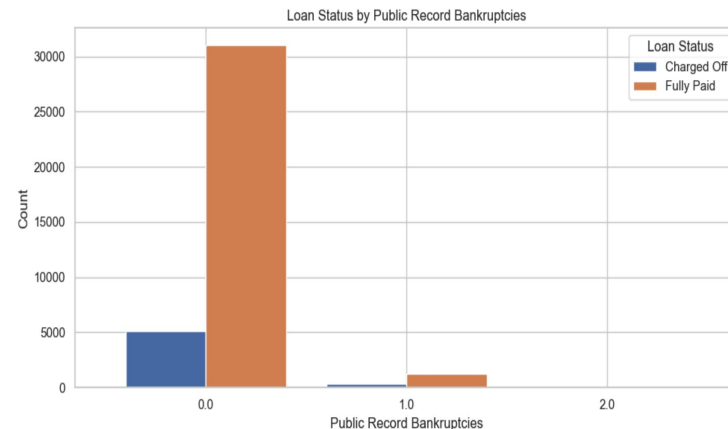


Fig 19.

Fig 18. The Public Record Bankruptcies Distribution Graph shows that most loan applicants have no public record bankruptcies, with only a small number having one or two. Bankruptcies are relatively rare among the applicants in this dataset.

Fig 19. The Loan Status by Public Record Bankruptcies Graph shows that borrowers with a higher number of public record bankruptcies have a lower rate of charge-offs. This suggests that factors such as credit recovery or rehabilitation efforts might be influencing outcomes for borrowers with bankruptcies.

Home Ownership Analysis and Segmentation

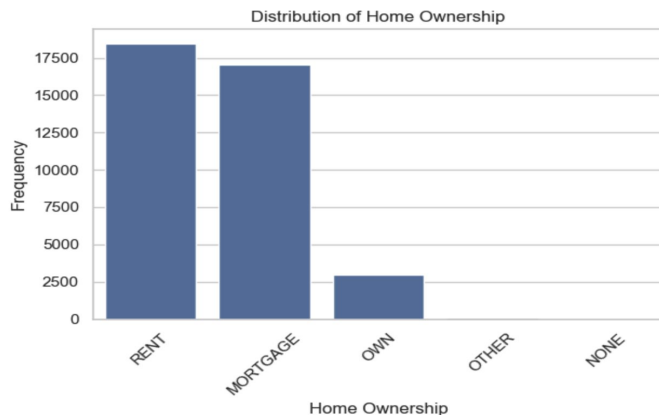


Fig 20.

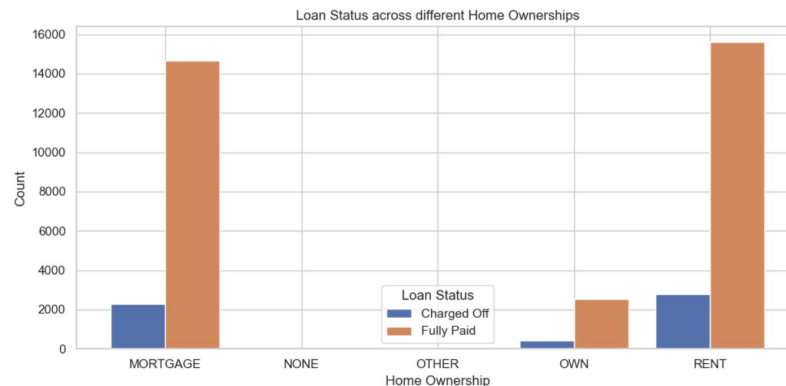


Fig 21.

Fig 20. The Home Ownership Distribution Graph shows that most borrowers are renters, followed by those with a mortgage. Fewer borrowers own their homes outright, and even fewer have other or no home ownership status.

Fig 21. The Loan Status across Different Home Ownership Graph shows that home ownership affects default rates. Borrowers who own their home have the lowest default rate, followed by those with a mortgage, while renters have significantly higher default rates.

Revolving Line Utilization Analysis and Segmentation

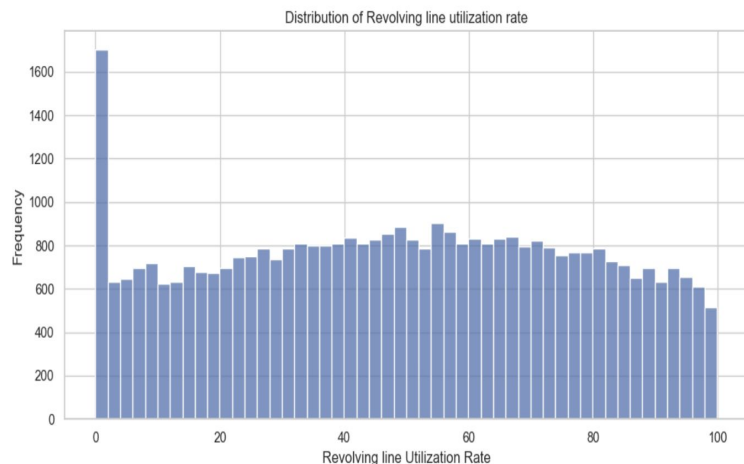


Fig 22.

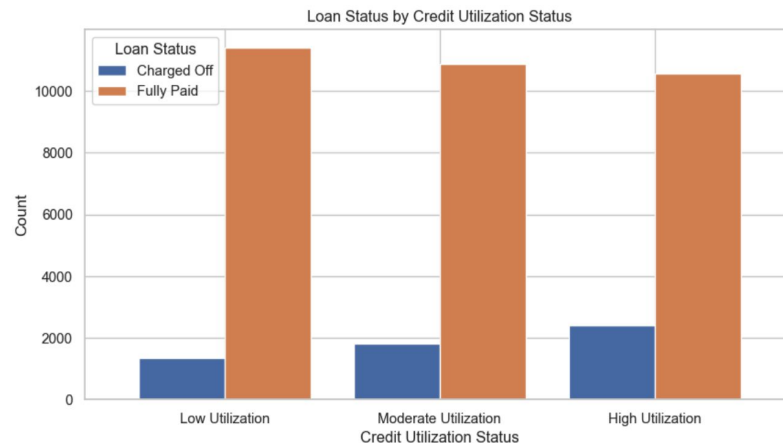


Fig 23.

Fig 22. The Revolving Line Utilization Rate Distribution Graph shows that most customers have a utilization rate of 0, indicating low credit line usage. The distribution is right-skewed, with a longer tail on higher utilization rates, suggesting that only a small proportion of customers have very high credit line usage.

Fig 23. Credit utilization status significantly impacts loan default rates. Borrowers with low credit utilization have the lowest default rates, followed by those with moderate utilization, while borrowers with high utilization show the highest default rates.

Multi-Variate Analysis

The multivariate analysis examines the interactions among multiple loan attributes, such as loan amount, interest rates, credit utilization, and annual income. This analysis reveals complex patterns and correlations that single-variable assessments may miss, providing a nuanced understanding of how these factors collectively influence loan outcomes. It enhances risk profiling and informs more precise decision-making strategies.

HeatMap Correlation for different continuous variable

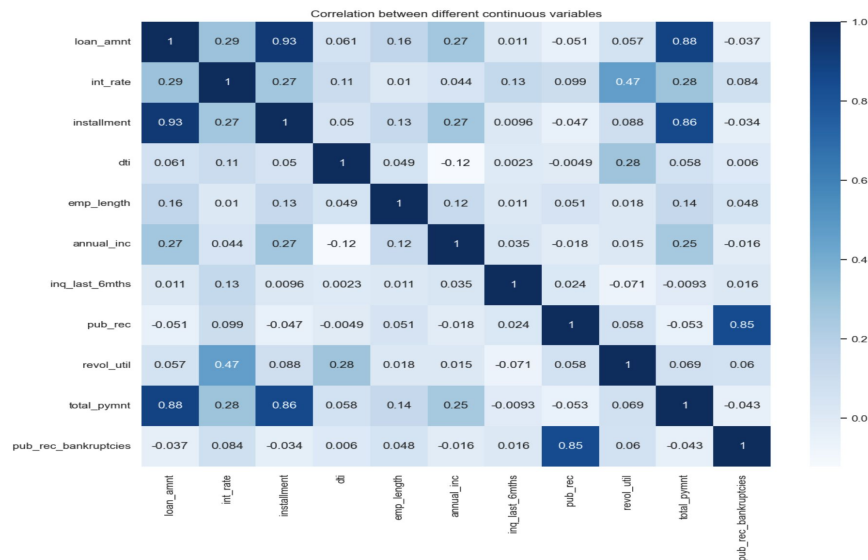


Fig 24.

Fig 24. The Correlation between Different Continuous Variables Graph highlights that larger loans lead to higher installments and total payments. Higher incomes are associated with larger loans, while higher interest rates increase installment amounts. Employment length has minimal impact on other financial variables. Higher income generally reduces the debt-to-income ratio. Additionally, higher revolving credit utilization tends to correlate with higher interest rates, and more public records are linked to a higher likelihood of bankruptcies.

Multivariate Analysis

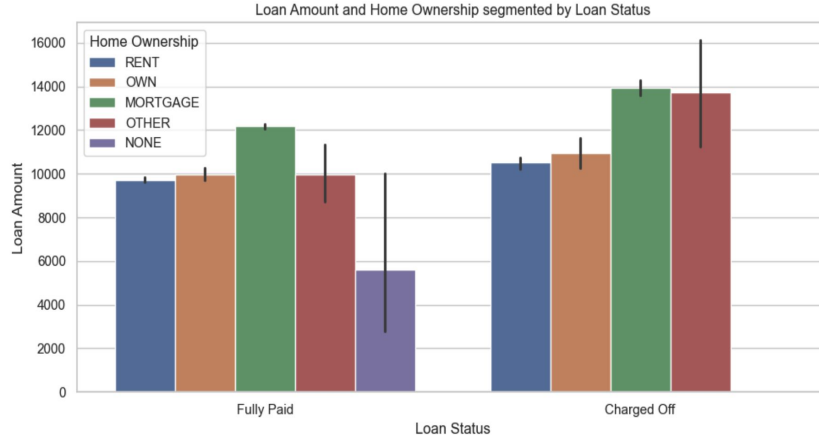


Fig 25.

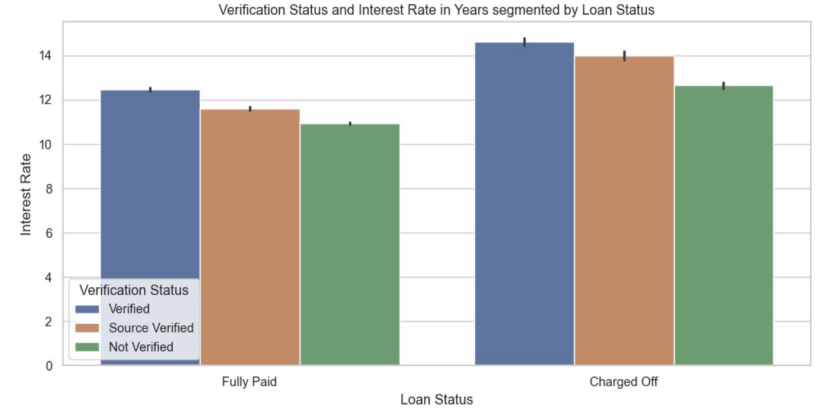


Fig 26.

Fig 25. The Loan Amount and Home Ownership Graph indicates that borrowers with higher loan amounts and stable homeownership statuses (e.g., 'MORTGAGE', 'OWN') are more likely to fully repay their loans. Those with 'RENT' or 'OTHER' are more prone to default. Loan amounts are generally higher for borrowers with a 'MORTGAGE' compared to those with 'RENT' or 'NONE'.

Fig 26. The Verification Status and Interest Rate Graph reveals that 'Verified' borrowers have slightly higher interest rates than 'Source Verified' or 'Not Verified' borrowers. 'Not Verified' borrowers have the highest rates, suggesting a link between verification status and default risk.

Multivariate Analysis

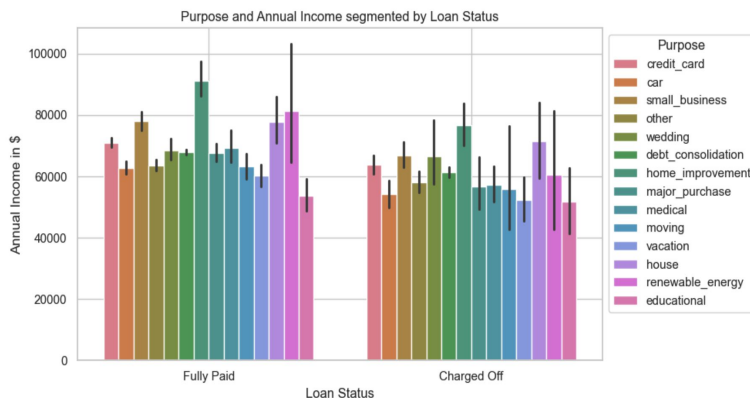


Fig 27.

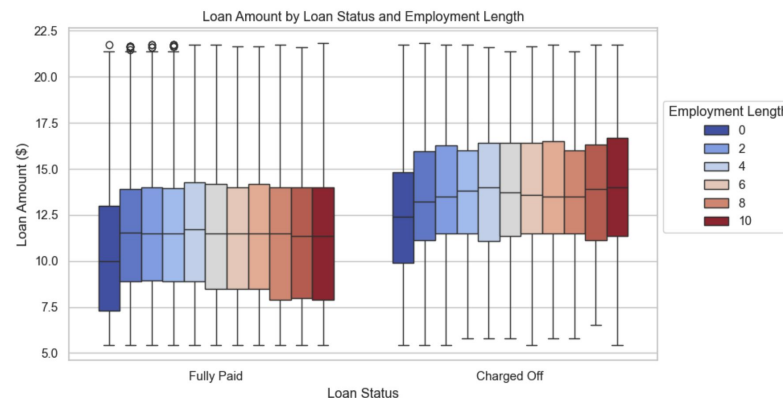


Fig 28.

Fig 27. The Purpose and Annual Income Graph shows that loans for 'credit_card' and 'car' have higher repayment rates, while 'small_business' and 'debt_consolidation' loans have lower rates. Higher annual income improves repayment rates, especially for 'home_improvement' and 'house' loans. Fully paid loans more commonly include 'credit_card' and 'car', while charged-off loans more often involve 'small_business' and 'debt_consolidation'.

Fig 28. The boxplot shows that interest rates for "Charged Off" loans are generally higher than those for "Fully Paid" loans, indicating a link between higher interest rates and a greater risk of default. Employment length also affects interest rates, as the varying widths and distributions across different employment durations suggest. Although a longer employment history may sometimes result in lower rates, this trend isn't consistent, indicating that other factors also influence interest rates.

Key Correlation

Key Correlations:

- **Loan Amount and Installment:** Larger loans lead to higher monthly payments.
- **Total Payment and Loan Amount:** Bigger loans result in higher total payment obligations.
- **Annual Income and Loan Amount:** Higher incomes are associated with larger loans.
- **Interest Rate and Installment:** Higher interest rates increase installment amounts.
- **Debt-to-Income Ratio and Annual Income:** Higher incomes generally mean lower debt-to-income ratios.
- **Revolving Utilization and Interest Rate:** Higher credit utilization correlates with higher interest rates.
- **Public Records and Bankruptcies:** More public records are linked to a higher likelihood of bankruptcy.

Loan Defaults:

- **Home Ownership:** Renters and those without homeownership default more often than homeowners.
- **Loan Purpose:** High-risk loans, like 'small_business' and 'debt_consolidation', have higher default rates.
- **Annual Income:** Lower-income borrowers are more prone to default.
- **Verification Status:** Loans to 'Not Verified' borrowers show higher default rates.
- **Employment Length:** Shorter employment tenures are linked to a higher risk of default.

Recommendations to Mitigate Charge-offs

Recommendations to Mitigate Charge-offs:

- **Improve Credit Scoring Models:** Factors such as revolving utilization, public records, and debt-to-income ratios should be better incorporated into this model for a more accurate risk-based credit evaluation.
- **Lend on Your Own Terms:** Offer loan terms like lower interest rates or a shorter repayment period to attract low-risk borrowers and incentivize quick payback, reducing your risk contributed.
- **Tighten Underwriting Criteria:** Make it harder to approve high risk loan purposes, particular 'small_business' and 'debt_consolidation' due to the higher default rates.
- **Financial Literacy Programs:** Offer necessary financial education to borrowers in order to improve their knowledge on debt management and responsible borrowing, which can drive default rates down the line.
- **Increase Debt Counseling Services:** Provide customized debt counseling to high-risk borrowers, give advice and tools on how managing debt properly and refrain from a default.
- **Practice Proactive Risk Management :** Roll out early warning systems to identify borrowers showing initial signs of financial difficulty so that timely remedial steps can be taken to prevent defaults.
- **Homeownership Promotion:** Support homeownership, which improves long-term financial security and reduces loan default probability.
- **Tighten Verification Process:** Solidify the verification processes by which a prospective borrower is scrutinized to lower the probability of giving out credit lines to borrowers with falsified information and therefore decrease default occurrences.

These strategies aim to improve risk management and borrower profiling, reducing the potential for charge-offs and defaults.